AD-A063 706   BROWN UNIV PROVIDENCE R I DIV OF APPLIED MATHEMATICS   F/G 12/1
APPROXIMATION OF SOLUTIONS TO DIFFERENTIAL EQUATIONS WITH RANDO--ETC(U)
NOV 78 H J KUSHNER   N00014-76-C-0279

UNCLASSIFIED.   NL

1 OF 1
AD
A063708

END
DATE
FILMED
3-79
DDC

6

APPROXIMATION OF SOLUTIONS TO DIFFERENTIAL EQUATIONS WITH RANDOM INPUTS BY DIFFUSION PROCESSES

12

10 Harold J. Kushner*
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

11 November ~~, 1978~~

15 N00014-76-C-0279, AFOSR-76-3063

12 22p.

Abstract

Let $y^{\varepsilon}(\cdot)$ denote a random process whose bandwidth, loosely speaking, goes to ∞ as $\varepsilon \to 0$. Consider the family of differential equations $\dot{x}^{\varepsilon} = g(x^{\varepsilon},y^{\varepsilon}) + f(x^{\varepsilon},y^{\varepsilon})/\alpha(\varepsilon)$, where $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$. The question of interest is: does the sequence $\{x^{\varepsilon}(\cdot)\}$ converge in some sense and if so which, if any, ordinary or Itô differential equation does it satisfy? Normally, the limit is taken in the sense of weak convergence. The problem is of great practical importance, for such questions arise in many practical situations arising in many fields. Often the limiting equation is nice and can be treated much more easily than can the $x^{\varepsilon}(\cdot)$. In any case, in practice approximations to properties of the $x^{\varepsilon}(\cdot)$ are usually sought in terms of ε and some limit. To illustrate these points, as well as a related stability problem, we give a practical example which arises in the theory of adaptive arrays of antennas.

The topic of convergence has seen much work, starting with the fundamental papers of Wong and Zakai, and followed by others, including Khazminskii, Papanicolaou and Kohler, etc. From a non-probabilistic point of view, it has been dealt with by McShane and Sussmann. In this paper, ~~we discuss~~ a rather general and efficient method of getting the correct limits is discussed. The idea exploits some general semigroup approximation results of Kurtz, and often not only gets better results than those obtained by preceding methods, but is also easier to use.

*This research was supported in part by the Air Force Office of Scientific Research under AF-AFOSR 76-3063, in part by the National Science Foundation under NSF-Eng 77-12946, and in part by the Office of Naval Research under N00014-76-C-0279-P0002.

1. Introduction

Let $y^\varepsilon(\cdot)$ denote a stationary random process whose "bandwidth" goes to ∞ as $\varepsilon \to 0$, and define the R^r-valued process $x^\varepsilon(\cdot)$ by the O.D.E.

$$(1.1) \qquad \dot{x}^\varepsilon = g(x^\varepsilon, y^\varepsilon) + f(x^\varepsilon, y^\varepsilon)/\alpha(\varepsilon), \quad x_0 = x(0) \text{ given,}$$

where $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$, and $Ef(x,y^\varepsilon) = 0$ for each x. In this paper, we address the question: as $\varepsilon \to 0$, what is the limit of $\{x^\varepsilon(\cdot)\}$; in particular, does it satisfy an ordinary or Itô stochastic differential equation, and if so, what is that equation. The question arises frequently in applications in many areas. Often $y^\varepsilon(\cdot)$ is a rather arbitrary process and yet the limit is a nice Markov process satisfying, say, an Itô equation. Then many functionals of $x^\varepsilon(\cdot)$ can be approximated by functionals of the limit and the parameter ε, for small ε. In applications, this is often done, either explicitly or implicitly. In Sections 2 and 6, one particular important application will be discussed.

The problem has been around for some time and is a crucial aspect of the problem of modelling the processes which arise in practice by mathematically tractable processes. Perhaps, the first mathematical treatment was given by Wong and Zakai [1], [2] who dealt with equations of the form

$$(1.2) \qquad \dot{x}^\varepsilon = g(x^\varepsilon) + f(x^\varepsilon)y^\varepsilon,$$

where $y^\varepsilon(\cdot)$ was (more or less) the derivative of a polygonal approximation $Y^\varepsilon(\cdot)$ to a Wiener process, and $Y^\varepsilon(\cdot)$ converged to that process as ε went to zero. Much subsequent was done on the following form. Let $\alpha(\varepsilon) = \varepsilon$, suppose that $y(\cdot)$ is a stationary bounded process and $\rho(\cdot)$ a measurable function which satisfy the strong mixing condition (1.3).

$$(1.3a) \qquad \int_0^\infty \rho^{1/2}(s)ds < \infty$$

$$(1.3b) \qquad |P\{B|A\} - P\{B\}| \le \rho(\tau)$$

for each t,τ and each $B \in \mathscr{B}(y_s, s \ge \tau + t)$ and $A \in \mathscr{B}(y_s, s \le t)$. Let $y^\varepsilon(t) = y(t/\varepsilon^2)$. Motivation for this scaling is given in the next subsection. (We write the values of a process $y(\cdot)$ as either $y(t)$ or y_t, depending on notational convenience.)

Under (1.3) and other conditions on g and f, (1.1) was dealt

78 12 27 087

with by Khazminskii [3], Papanicolaou [5], Papanicolaou and Kohler [4], Papanicolaou and Blankenship [6] and Kushner [7]. The last reference obtained perhaps the most general results (for the time invariant case) and allowed cases where $y^{\varepsilon}(\cdot)$ could contain (approximations to) impulsive jumps, and also where $y(\cdot)$ is unbounded but the form $f(x,y) = f(x)y$ was used.

Let C_0^i denote the space of real-valued functions on R^r which go to zero as $|x| \to \infty$, together with their first i^{th} mixed partial derivatives, and let $\hat{C}^i$ denote the subset with compact support. Let subscript x denote gradient, and define the operator A on $\hat{C}^2$ by

$$(1.4)\quad Ak(x) = Eg'(x,y_s)k_x(x) + \int_0^{\infty} Ef'(x,y_s)(f'(x,y_{s+\tau})k_x(x))_x d\tau$$
$$\equiv \sum_i b_i(x)\frac{\partial k(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 k(x)}{\partial x_i \partial x_j}.$$

In references [3]-[7], it was proved (under various conditions on $f,g,y(\cdot)$) that $x^{\varepsilon}(\cdot)$ converged weakly to a diffusion process $x(\cdot)$ whose generator on $\hat{C}^2$ functions is the operator A of (1.4).

References [4] and [6] contain a wealth of ideas on the approximation and related problems. The methods used in [7] are based on general semigroup approximation methods of Kurtz [8]. They have a number of advantages over previous methods, being somewhat easier to use and giving better results in many cases. The method will be described and used in Section 3.

Most past work has dealt with showing that $x(\cdot)$ is a good approximation to $x^{\varepsilon}(\cdot)$ in some sense. Only recently (see, e.g. [13]) has the question of dealing with the control and stability properties of $x^{\varepsilon}(\cdot)$ in terms of those of $x(\cdot)$ been considered. Reference [14] deals with the reversed problem: finding $x^{\varepsilon}(\cdot)$ which are easier to work with than $x(\cdot)$.

<u>Discussion of properties of $y^{\varepsilon}(\cdot)$ as $\varepsilon \to 0$</u>. As $\varepsilon \to 0$ in (1.1) the process $f(x^{\varepsilon},y^{\varepsilon})/\alpha(\varepsilon)$ is "increasingly compressed", hence (loosely speaking) the bandwidth goes to ∞. If $f(x,y)$ were not divided by $\alpha(\varepsilon)$, then the average energy in the process $f(x_t^{\varepsilon},y_t^{\varepsilon})/\alpha(\varepsilon)$ (over any finite interval) would tend to zero as $\varepsilon \to 0$ and the f term would play no role in the limit. To see the rough idea most simply, let $y^{\varepsilon}(\cdot)$ be scalar valued, let $f(x,y) = f(x)y$, let $R(\cdot)$ and $S(\cdot)$ denote the correlation function and spectral density of a

stationary process $y(\cdot)$, consider the special case where $y_t^\varepsilon = y(t/\varepsilon^2)$ and let $R^\varepsilon(\cdot)$ and $S^\varepsilon(\cdot)$ denote the correlation function and spectral density of $y^\varepsilon(\cdot)$. Then $R^\varepsilon(t) = R(t/\varepsilon^2)/\alpha^2(\varepsilon)$ and

$$S^\varepsilon(w) = \int_{-\infty}^{\infty} e^{iwt} R^\varepsilon(t)dt = \int_{-\infty}^{\infty} e^{iwt} R(t/\varepsilon^2)dt/\alpha^2(\varepsilon) = \varepsilon^2 S(\varepsilon^2 w)/\alpha^2(\varepsilon).$$

Unless $\alpha(\varepsilon) = \varepsilon$, the energy per unit bandwith either blows up $(\varepsilon/\alpha(\varepsilon) \to \infty)$ or goes to zero $(\varepsilon/\alpha(\varepsilon) \to 0)$. When $\alpha(\varepsilon) = \varepsilon$, $R^\varepsilon(0) = R(0)/\varepsilon^2 \to \infty$. If the "magnitude" of $y^\varepsilon(t)/\varepsilon$ did not go to ∞ as $\varepsilon \to \infty$, then the energy per unit bandwidth would go to zero. So, in order to get a constant energy per unit bandwidth as $\varepsilon \to 0$, we need both a time compression (t/ε^2 scale) and an amplitude magnitication ($\alpha(\varepsilon) = \varepsilon$). Use of this remark will be made in the next section.

In Section 5, we illustrate the technique of [7] on an important class of problems not explicitly treated previously. For each ε, let $\{s_i^\varepsilon, i \geq 0\}$, denote a stationary process. Define $s^\varepsilon(\cdot)$ as the function which is equal to s_i^ε on the interval $[i,i+1)$, set $\xi_t^\varepsilon = s^\varepsilon(t/\varepsilon)$, and let s_i^ε be "small"; i.e., $Ef(x,s_i^\varepsilon) = 0$, $\text{var } f(x,s_i^\varepsilon) \approx \varepsilon$ and define $x^\varepsilon(\cdot)$ by

$$\dot{x}^\varepsilon = g(x^\varepsilon,\xi^\varepsilon) + f(x^\varepsilon,\xi^\varepsilon)/\varepsilon, \quad x(0) = x_0. \tag{1.5}$$

The exact forms of the conditions to be used are stated in Section 5. The form (1.5) is chosen partly to illustrate the method. That $\xi^\varepsilon(\cdot)$ is piecewise constant makes the calculations a little easier, but is not a particularly crucial assumption. We will treat the case where $f(\cdot,\cdot)$ is linear in its second argument: $f(x,s) = f(x)s$.

Equation (1.5) is also important from the point of view of applications. Consider the scalar valued discrete parameter sequence $X_{n+1}^\varepsilon = X_n^\varepsilon + h(X_n^\varepsilon,s_n^\varepsilon)$ where $Eh(x,s_i^\varepsilon) \equiv \varepsilon p(x)$ and $\text{var } h(x,s_i^\varepsilon) = \sigma^2(x)\varepsilon$. Then, setting $q(x,s_i^\varepsilon) = h(x,s_i^\varepsilon) - Eh(x,s_i^\varepsilon)$ yields the discrete parameter version of (1.5):

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon p(X_n^\varepsilon) + q(X_n^\varepsilon,s_n^\varepsilon). \tag{1.6}$$

Let $\bar{x}^\varepsilon(\cdot)$ denote a piecewise linear interpolation of $\{X_n^\varepsilon\}$ which is linear in each $[\varepsilon n,\varepsilon n+\varepsilon)$ and equals X_n^ε at εn. Then the slope of $\bar{x}^\varepsilon(\cdot)$ is $\rho(X_n^\varepsilon) + q(X_n^\varepsilon,s^\varepsilon(\varepsilon n/\varepsilon))/\varepsilon$ in $[\varepsilon n,\varepsilon n+\varepsilon)$. Thus, (1.5) is a continuous parameter version of $\{X_n^\varepsilon\}$. The limits of $\bar{x}^\varepsilon(\cdot)$ and of $\{x^\varepsilon(\cdot)\}$ are not necessarily the same, although in many cases we can find g and g such that $x^\varepsilon(n\varepsilon) = X_n^\varepsilon$ for all ε, n.

Let us suppose that in (1.1), $\alpha(\varepsilon) = \varepsilon$ and $y^\varepsilon(t) = y(t/\varepsilon^2)$. Equation (1.5) differs from equation (1.1) in that the $\xi^\varepsilon(t)$ essentially become small in some sense as $\varepsilon \to 0$. But the scaling is also different (less compression), t/ε being used in lieu of t/ε^2. Equation (1.5) (and (1.6)) correspond to a problem where, as $\varepsilon \to 0$, more and more random effects affect the system, but where the individual effects became smaller and smaller. Let $f(x,s) = f(x)s$ and write $\xi^\varepsilon(t)/\varepsilon = [s^\varepsilon(t/\varepsilon)/\sqrt{\varepsilon}]/\sqrt{\varepsilon}$, bringing the form (1.5) into that of (1.1) but with $\sqrt{\varepsilon}$ replacing ε and $s^\varepsilon(t/\varepsilon)/\sqrt{\varepsilon}$ replacing $y^\varepsilon(t)$. But now, as $\varepsilon \to 0$, $s^\varepsilon(t/\varepsilon)/\sqrt{\varepsilon}$ might become unbounded. Owing to this, the methods used for (1.1) (at least when $y(\cdot)$ was assumed bounded) need to be modified a little for use here.

In Section 2, we discuss a currently important problem concerning "adaptive" antenna arrays, which illustrates one particular value derived from the type of limit results with which we are concerned. Sections 3 and 4 describe Kurtz's [8] interesting method for proving tightness and weak convergence of a sequence of not necessarily Markov processes. Henceforth, x^ε is used only for the solution to (1.5). Convergence of the finite dimensional distributions of $x^\varepsilon(\cdot)$ to those of a particular diffusion $x(\cdot)$ is proved in Section 5. Also, Section 5 proves tightness of $\{X^\varepsilon(\cdot)\}$. Together these results yield that $x^\varepsilon(\cdot)$ converges weakly to $x(\cdot)$. In Section 6, we return to the antenna problem, and treat the problem of weak convergence and get a moment estimate for the adapting parameters, and discuss a related stability problem.

The use of weak convergence methods seems quite natural for our problem. Often w.p.1 results are meaningless, since usually only one system (a fixed ε) is to be studied, and we seek approximations to its properties in terms of ε and properties of the limit.

2. A Problem in Adaptive Antenna Arrays

Let $n(\cdot) = (n_1(\cdot), \ldots, n_r(\cdot))$ denote a "wide band" complex valued stochastic process. We are given an array of r antennas with received signal plus noise $v(t) = s(t) + n(t) = \{s_i(t) + n_i(t)\}$, $s_i(t)$ and $n_i(t)$ being complex valued. The $v(t)$ is multiplied by a complex valued weight w, and the object is to find the weights which maximize the ratio of signal to noise power in the output $w'v(t)$. The problem is important and of great current interest (see the papers in [9] or [10] and references contained therein). The signal frequency is known, the signals received by the antennas differ only in the phase. Let * denote complex conjugate. Let $S_0^* = (1, \exp i\phi_2, \ldots, \exp i\phi_r)$, where ϕ_j is the phase of $s_j(t)$ relative to that of $s_1(t)$, and let S be proportional to S_0. With $\bar{M} = En^*(t)n'(t)$, the optimum weight is $w = k\bar{M}^{-1}S^*$, for any constant $k > 0$.

In many applications, $\bar{M}$ is time varying, due to deliberate jamming attempts, or due to more natural phenomena. In fact, in many applications $n(\cdot)$ is a strong competing signal which we wish to "tune out" and its covariance may vary, depending on the particular use to which the system is put. We suppose (as is often the case - e.g., in pulsed radar) that the signal power is much less than the noise power, so that $\bar{M} \approx Ev^*(t)v'(t)$.

A very useful and relatively simple mechanism for adapting the weights (see, e.g. [10]) can be constructed. The relevant equation is $(M_t = v^*(t)v'(t))$

$$\tau\dot{w} + (GM+I)w = G_0S, \tag{2.1}$$

where τ is a scalar system time constant and G and G_0 are system gains. Since M is the "square" of a wide band process, if the bandwidth (BW) goes to infinity and the energy per unit BW does not go to zero, (2.1) becomes meaningless. In practice, we are interested in both Ew_t and in an equation for an approximation to $w_t - Ew_t$ for wide BW noise.

A commonly used "engineering" heuristic argument says that since $M(\cdot)$ is wide band and $w(\cdot)$ is much smoother than $M(\cdot)$, the two are essentially independent and $EM(t)w(t) \approx EM(t)Ew(t)$ and that Ew_t approximately equals $\bar{w}_t$, the solution to

$$\dot{\bar{w}} + (G\bar{M}+I)\bar{w} = G_0S^*. \tag{2.2}$$

Of course (2.2) does not give the correct value of Ew_t, even as an approximation, unless the energy per unit BW of the noise is very small. To see this, simply consider the scalar case where $\tau\dot{w} + (Gn^2+1)w = G_0$; solve it and take expectations. Since (2.2) is widely used, we must find an interpretation with respect to which it makes sense. If (2.2) is an asymptotic result, then it must be satisfied by a limit of solutions to (2.1) (or their expectations), as some parameter tends to say, ∞. The comments below are illustrative of the usefulness of the limit results to which this paper (and references [3]-[7]) are devoted.

Let σ^2 denote the sum of the eigenvalues of $\bar{M}$. In practice, often a rough estimate of a quantity proportional to σ^2 (the noise power) is made, and an automatic gain control mechanism is used to adjust G, usually decreasing G as the estimate of σ^2 increases. Such a mechanism is crucial to the proper scaling of (2.1), and we assume its use. In fact, suppose that for some number K, $G = K/\sigma^2$. Then (approximately, actually, since we ignore the "signal" component of M) with $\delta M = M - \bar{M}$,

$$\tau\dot{w} + [K\bar{M}/\sigma^2 + K(\delta M)/\sigma^2 + I]w = G_0S^*. \tag{2.3}$$

As the BW of $n(\cdot)$ tends to ∞, the effects of $K(\delta M)w/\sigma^2$ become negligable (a consequence of the type of argument in [7], Sections 6 and 7, under reasonable assumptions on $n(\cdot)$) and the limit is precisely the solution of (2.2). For concreteness, we consider the case arising from (1.1) where $n_t = y(t/\varepsilon^2)/\varepsilon$ and $y(\cdot)$ is a stationary bounded process and $\varepsilon = 1/\sigma$. Set $\bar{M}_0 = Ey_t^*y_t'$, $\delta\tilde{M}_t^\varepsilon = [y^*(t/\varepsilon^2)y'(t/\varepsilon^2) - \bar{M}_0]$. Write δM_t as $\delta M_t^\varepsilon \equiv \delta\tilde{M}_t^\varepsilon/\varepsilon^2$ and use $\delta\tilde{M}_t$ for $\delta\tilde{M}_t^1$.

Now, define $\delta w_t = w_t - \bar{w}_t$ and $u^\varepsilon = \sigma\delta w$. Then

$$\begin{aligned} &\tau\dot{u}^\varepsilon + [K\bar{M}/\sigma^2 + I]u^\varepsilon + K(\delta M^\varepsilon/\sigma^2)u^\varepsilon + K(\delta M^\varepsilon/\sigma)\bar{w} = 0, \\ &\dot{u}^\varepsilon = -\frac{1}{\tau}[K\bar{M}_0 + I]u^\varepsilon - \frac{K}{\tau}(\delta\tilde{M}^\varepsilon)u^\varepsilon - \frac{K}{\tau}\left(\frac{\delta\tilde{M}^\varepsilon}{\varepsilon}\right)\bar{w}. \end{aligned} \tag{2.4}$$

As BW $\to \infty$, the effects of $K(\delta M^\varepsilon/\sigma^2)u$ disappear and $K(\delta M^\varepsilon/\sigma)\bar{w}$ becomes "white noise", in the sense that there is a standard Wiener process $B(\cdot)$ such that the limit process has the law of $u(\cdot)$ in

$$\tau du + [K\bar{M}_0 + I]u\,dt + Q\,dB = 0, \quad u(0) = 0. \tag{2.5}$$

Q is obtainable by the method of Theorem 5. We return to this problem in Section 6, and deal with the convergence problem and a related stability problem when all quantities are <u>not</u> complex valued, to simplify the notation.

3. Convergence of Finite Dimensional Distributions

In reference [8], Kurtz gave some fairly general methods for showing convergence to a Markov process of a sequence of non-Markov processes, either in the sense of weak convergence or in the sense of convergence of finite dimensional distributions. In this section and in the next, we briefly describe his method. Later we apply it, together with an idea in [5], [6], to get limit results in a fairly efficient manner.

Sections 3 and 4 are identical to Sections 2 and 3 of [7].

Let $(\Omega, P, \mathscr{F})$ denote a probability space, $\{\mathscr{F}_t\}$ a nondecreasing sequence of sub σ-algebras of $\mathscr{F}$, let $\mathscr{L}$ denote the space of progessively measurable real valued processes $k(\cdot)$ on $[0,\infty)$, adapted to $\{\mathscr{F}_t\}$ and such that $\sup_t E|k(t)| < \infty$. Let k^ε and k be in $\mathscr{L}$. Define the limit "p-lim" by p-lim $k^\varepsilon = k$ iff $\sup_{\varepsilon>0} \sup_t E|k^\varepsilon(t)| < \infty$ and $E|k^\varepsilon(t) - k(t)| \to 0$ for each t as $\varepsilon \to 0$. For each $s > 0$, define the operator $\mathscr{T}(s): \mathscr{L} \to \mathscr{L}$ by $\mathscr{T}(s)k$ = function in $\mathscr{L}$ whose value at t is the random variable $E_{\mathscr{F}_t} k(t+s)$. There is a version which is progressively measurable ([8], Appendix) and we always assume that this is the one which is used. The $\hat{\mathscr{T}}(s)$, $s \geq 0$, are a semigroup of linear operators on $\mathscr{L}$. Let $\hat{\mathscr{L}}_0$ denote the subspace of p-right continuous functions. If the limit $\text{p-lim}_{s\to 0} [\frac{1}{s}(\mathscr{T}(s)k-k)]$ and exists and is in $\hat{\mathscr{L}}_0$, we call it $\hat{A}k$ and say that $k \in \mathscr{D}(\hat{A})$. The operators $\mathscr{F}(s)$ and $\hat{A}$ are analogous to the semigroup and weak infinitesimal operator of a Markov process. Among the properties to be used later is ([8], equation (1.9))

$$\mathscr{T}(s)k - k = \int_0^s \mathscr{T}(\tau)\hat{A}k\,d\tau, \quad k \in \mathscr{D}(\hat{A}), \tag{3.1a}$$

or, equivalently,

$$E_{\mathscr{F}_t} k(t+s) - k(t) = \int_0^s E_{\mathscr{F}_t} \hat{A}k(t+\tau)\,d\tau, \quad \text{for each } t \geq 0. \tag{3.1b}$$

If, for some process $Z^\varepsilon(\cdot)$, $\mathscr{F}_t = \mathscr{B}(Z_s^\varepsilon, s \leq t)$, we may write $\mathscr{F}_t^\varepsilon, T_t^\varepsilon$ and $\hat{A}^\varepsilon$ for $\mathscr{F}_t$, $\mathscr{T}(t)$ and $\hat{A}$, resp.

The following Theorem (a specialization of [8], Theorem 3.11) is our main tool for dealing with (1.1) or (1.5).

<u>Theorem 1</u>. <u>Let</u> $Z^\varepsilon(\cdot) = x^\varepsilon(\cdot), \xi^\varepsilon(\cdot)$, $\varepsilon > 0$, <u>denote a sequence of</u> $R^{r+r'}$ <u>valued right continuous processes</u>, $x(\cdot)$ <u>a</u> (R^r-<u>valued</u>) <u>Markov process with semigroup</u> T_t <u>mapping</u> C_0 <u>into</u> C_0 <u>and which is strongly continuous</u> (<u>sup norm</u>) <u>on</u> C_0. <u>For some</u> $\lambda > 0$ <u>and dense set</u> D <u>in</u> C_0, <u>let</u> Range $(\lambda - A|_D)$ <u>be dense in</u> C_0 (<u>sup norm</u>, A = <u>infinitesimal operator of</u> $x(\cdot)$). <u>Suppose that, for each</u> $k \in D$, <u>there is a sequence</u> $\{k^\varepsilon\}$ <u>of progressively measurable functions adapted to</u> $\{\mathscr{F}_t^\varepsilon\}$ <u>and such that</u>

$$\text{p-lim}[k^\varepsilon - k(x^\varepsilon(\cdot))] = 0 \tag{3.2}$$

(3.3) $$\text{p-lim}[\hat{A}k^{\varepsilon} - Ak(x^{\varepsilon}(\cdot))] = 0.$$

Then, if $x^{\varepsilon}(0) \to x(0)$ weakly, the finite dimensional distributions of $x^{\varepsilon}(\cdot)$ converge to those of $x(\cdot)$.

Equations (3.2) and (3.3) are equivalent to (the limits are taken for each t as $\varepsilon \to 0$)

(3.2') $$\sup_{\varepsilon,t} E|k^{\varepsilon}(t) - k(x^{\varepsilon}(t))| < \infty,\ E|k^{\varepsilon}(t) - k(x^{\varepsilon}(t))| \to 0$$

(3.3') $$\sup_{\varepsilon,t} E|\hat{A}^{\varepsilon}k^{\varepsilon}(t) - Ak(x^{\varepsilon}(t))| < \infty,\ E|\hat{A}^{\varepsilon}k^{\varepsilon}(t) - Ak(x^{\varepsilon}(t))| \to 0.$$

4. Tightness

Let $\xi^{\varepsilon}(\cdot), x^{\varepsilon}(\cdot)$ denote the functions in the model (1.5). Let $\mathcal{F}_t^{\varepsilon}$ denote $\mathcal{B}(\xi_u^{\varepsilon}, u \le t)$ and write E_t^{ε} for $E_{\mathcal{F}_t^{\varepsilon}}$.

Again, we describe results from [8]. Let $D^r[0,\infty)$ denote the space of R^r valued functions on $[0,\infty)$ which are right continuous on $[0,\infty)$ and have left hand limits on $(0,\infty)$. Note that $x^{\varepsilon}(\cdot) \in D^r[0,\infty)$ w.p.1. Suppose that the finite dimensional distributions of $x^{\varepsilon}(\cdot)$ converge to those of a process $x(\cdot)$, where $x(\cdot)$ has paths in $D^r[0,\infty)$ w.p.1. Then, as noted in [8], bottom of page 628, $\{x^{\varepsilon}(\cdot)\}$ is tight in $D^r[0,\infty)$ if $\{k(x^{\varepsilon}(\cdot))\}$ is tight in $D[0,\infty)$ for each $k \in \hat{C}$. ($\hat{C}$ is used there, but it can be replaced by $\hat{C}^3$ or by any set of functions dense in $\hat{C}$ in the sup norm.) It follows from [8], Theorem 4.20, that $\{k(x^{\varepsilon}(\cdot))\}$ is tight in $D[0,\infty)$ if $x_0^{\varepsilon} \to x_0$ weakly and if, for each real $T > 0$, there is a random variable $\gamma_{\varepsilon}(\delta)$ such that

(4.1) $$E_t^{\varepsilon}\gamma_{\varepsilon}(\delta) \ge E_t^{\varepsilon} \min\{1, [k(x_{t+u}^{\varepsilon}) - k(x_t^{\varepsilon})]^2\},$$

for all $0 \le t \le T$, $0 \le u \le \delta \le 1$, and

(4.2) $$\lim_{\delta\to 0} \overline{\lim_{\varepsilon\to 0}} E\, \gamma_{\varepsilon}(\delta) = 0.$$

In [8], p. 629, Kurtz suggests a method of getting the $\gamma_{\varepsilon}(\delta)$. This method is developed in Theorem 2 and is used in the sequel. The k^{ε} below will be obtained in the same manner as we will obtain the k^{ε} needed in Theorem 1. We have $(||k|| = \sup_x |k(x)|)$

(4.3) $$E_t^\varepsilon[k(x_{t+u}^\varepsilon) - k(x_t^\varepsilon)]^2 \leq 2||k||\ |E_t^\varepsilon k(x_{t+u}^\varepsilon) - k(x_t^\varepsilon)| + |E_t^\varepsilon k^2(x_{t+u}^\varepsilon) - k^2(x_t^\varepsilon)|.$$

Theorem 2. Let $k \in \hat{C}^3$, and let there be a sequence $\{k^\varepsilon\}$ in $\mathscr{L}$, where $(k^\varepsilon)^i \in \mathscr{D}(\hat{\Lambda}^\varepsilon)$, $i = 1,2$, and such that, for each real $T > 0$ there is a random variable M^ε such that

(4.4) $$\sup_{t \leq T}|k^\varepsilon(t) - k(x_t^\varepsilon)| \to 0 \quad \text{in probability as} \quad \varepsilon \to 0,$$

(4.5) $$\sup_{t \leq T}|\Lambda^\varepsilon(k^\varepsilon(t))^i| \leq M_\varepsilon, \quad i = 1,2, \quad \lim_N \sup_\varepsilon P\{M_\varepsilon > N\} = 0.$$

Then $\{k(x^\varepsilon(\cdot))\}$ is tight in $D[0,\infty)$.

Proof: We need only show tightness of $\{k^\varepsilon(\cdot)\}$ on $[0,T]$. By (3.1)

$$E_t^\varepsilon(k^\varepsilon(t+u))^i - (k^\varepsilon(t))^i = \int_0^u E_t^\varepsilon \hat{\Lambda}^\varepsilon(k^\varepsilon(t+\tau))^i d\tau,$$

from which (4.3) and (4.5) yield that there is a $\gamma_\varepsilon(\delta)$ satisfying (4.1) and (4.2) for $\{k^\varepsilon(\cdot)\}$. Then (4.4) and tightness of $\{k^\varepsilon(\cdot)\}$ imply tightness of $\{k(x^\varepsilon(\cdot))\}$. Q.E.D.

5. Convergence of the Sequence $\{x^\varepsilon(\cdot)\}$ of (1.5)

We follow the general line of development in [7], using the ideas in Sections 3 and 4.

Assumptions:

(A1) Let $f(x,s) = f(x)s$. The functions $g(\cdot,\cdot)$ and $f(\cdot)$ are continuous, the first (second, resp.) function having continuous first (second, resp.) mixed partial x-derivatives.

(A2) There is a constant K such that

$$|f(x)| + |g(x,s)| \leq K(1+|x|).$$

(A3) $\{s_i^\varepsilon\}$ is bounded (uniformly in ε) and stationary and $Es_i^\varepsilon \equiv 0$.

Define $\mathscr{F}_i^\varepsilon = \mathscr{B}(s_j^\varepsilon,\ j \leq i)$, $\mathscr{F}_t^\varepsilon = \mathscr{B}(\xi_s^\varepsilon,\ s \leq t)$ and let E_i^ε and E_t^ε denote the corresponding conditional expectation operators.

Throughout the sequel, K denotes a constant, whose value may change from usage to usage. Also in (A4b,c), we let s_i^ε denote an arbitrary scalar component of itself. It seems to be most convenient to use (A4) in its given form. Other forms, more closely resembling strong mixing can be given. Strong mixing would imply the second part of (A4a) and (A4b,d). But something like (A4c) would still be required, because we need the ε dependence there.

(A4) a. <u>There are</u> μ_j <u>and</u> $\delta > 0$ <u>such that</u> $\sum_j \mu_j^{1/2} < \infty$ <u>and</u>

$$E|s_i^\varepsilon|^2 \le K\varepsilon, \quad |E_i^\varepsilon s_{i+j}^\varepsilon| \le K\mu_j.$$

b. $$|E_i^\varepsilon s_{i+j}^\varepsilon s_{i+j+k}^\varepsilon - Es_{i+j}^\varepsilon s_{i+j+k}^\varepsilon| \le K\mu_j$$

c. $$E^{1/2}|E_i^\varepsilon s_{i+j}^\varepsilon s_{i+j+k}^\varepsilon - Es_{i+j}^\varepsilon s_{i+j+k}^\varepsilon|^2 \le K\mu_j \varepsilon^{1/2+\delta}$$

d. $$\sum_j |E_i^\varepsilon g(x,s_{i+j}^\varepsilon) - Eg(x,s_{i+j}^\varepsilon)| < \infty,$$

$$\sum_j |E_i^\varepsilon g_x(x,s_{i+j}^\varepsilon) - Eg_x(x,s_{i+j}^\varepsilon)| < \infty.$$

<u>Let the expectation of the sums in</u> (d) <u>go to zero as</u> $\varepsilon \to 0$, <u>uniformly on bounded</u> x-<u>sets</u>.

Define the operators A_t^ε and $\overline{A}^\varepsilon$ by $(k(\cdot) \in \hat{C}^2)$

$$A_t^\varepsilon k(x) = E\int_0^\infty f'(x,\xi_t^\varepsilon)[k_x'(x)f(x,\xi_{t+u}^\varepsilon)]_x du/\varepsilon^2 \tag{5.1}$$

$$= E\int_0^\infty f'(x,s^\varepsilon(t/\varepsilon))[k_x'(x)f(x,s^\varepsilon(u+t/\varepsilon))]_x du/\varepsilon$$

and

$$\overline{A}^\varepsilon k(x) = \lim_{N\to\infty}\frac{1}{N}\int_0^N A_{t+s}^\varepsilon k(x)ds.$$

Let $i = [t/\varepsilon]$, which equals the nearest integer to t/ε which is not larger than t/ε. Then

$$A_t^\varepsilon k(x) = \frac{1}{\varepsilon}\sum_{j=1}^\infty Ef'(x,s_0^\varepsilon)[k_x'(x)f(x,s_j^\varepsilon)]_x \tag{5.2}$$

$$+ \frac{1}{\varepsilon} Ef'(x,s_0^\varepsilon)[k_x'(x)f(x,s_0^\varepsilon)]_x(\varepsilon[t/\varepsilon] + \varepsilon - t)/\varepsilon.$$

Note that $\bar{A}^{\varepsilon}$ is A_t^{ε} but with $(\varepsilon[t/\varepsilon] + \varepsilon - t)/\varepsilon$ replaced by 1/2.

(A5) The sum in (5.2) converges uniformly in ε for each x and $k(\cdot)$. Hence also uniformly in x, since $f(x,s) = f(x)s$. There are matrix and vector valued functions $a(\cdot)$ and $b(\cdot)$, resp., such that $(\bar{g}^{\varepsilon}(x) = Eg(x,s_i^{\varepsilon}))$

$$\bar{A}^{\varepsilon}k(x) + k_x'(x)\bar{g}^{\varepsilon}(x) \to b'(x)k_x(x) + \frac{1}{2}\sum_{i,j} a_{ij}(x)\partial^2 k(x)/\partial x_i \partial x_j \equiv Ak(x) \tag{5.3}$$

uniformly in x for each k.

(A6) A is the restriction to $\hat{C}^2$ of the strong infinitesimal operator of a strong Markov diffusion process (with no finite escape time) with semigroup T_t, which maps C_0 into C_0 and is strongly continuous on C_0.

(A7) For some real $\lambda > 0$, the set $(\lambda - A)\hat{C}^2$ is dense in C_0.

We note that it is enough to use $\hat{C}^3$ in (A5) and (A6).

Remark on bounding the coefficients. In our case (since $f(x,s) = f(x)s$) there is a matrix $\sigma(\cdot)$ such that $a(\cdot) = \sigma(\cdot)\sigma'(\cdot)$. Suppose that b,σ, are locally Lipschitz and that $x(\cdot)$, the diffusion with coefficients $\sigma(\cdot),b(\cdot)$, has no finite escape time w.p.1. Then $x(\cdot)$ has a representation as a solution to an Itô equation, and can be defined on Wiener process space. We say that $x^N(\cdot)$ is an N-truncation of $x(\cdot)$ if its drift and diffusion coefficients $b^N(\cdot),\sigma^N(\cdot)$ are bounded and equal $b(\cdot),\sigma(\cdot)$ in $S_N = \{x: |x| \le N\}$, and are obtained by a bounding of g and f. The process $x^N(\cdot)$ can be defined on the same Wiener process space on which $x(\cdot)$ is defined, and then equals $x(\cdot)$ up until the first exit time from S_N (which $\to \infty$ w.p.1 as $N \to \infty$). If for each N, Theorem 2 (and, of course, (A6)-(A7)) is true for some N-truncation, then it is true as stated.

Remark on the conditions. (A6) and (A7) are required for the use of (the semigroup approximation) Theorem 1. (A7) is equivalent to the strong infinitesimal operator of the T_t of (A6) being the closure of the operator A of (5.3) acting on $\hat{C}^2$ (or on any set sense in $\hat{C}^2$ in the norm $||k||_2 = \sup_x(|k(x)| + |k_x(x)| + |k_{xx}(x)|)$).

This condition does not seem to be very restrictive. As noted above it is often only necessary to verify it for some $x^N(\cdot)$ truncation for each N. If the coefficients are bounded and uniformly twice continuously differentiable, then ([11], Chapter 8.4) T_t maps C_0^2 into C_0^2 and is strongly continuous on C_0^2 with respect to norm $||k||_2$. We can then replace C_0 by C_0^2 in (A5), (A6) and in Theorem 1 and consider T_t as acting on C_0^2 rather than on C_0.

In Theorem 3, we apply Theorem 1 to get convergence of finite dimensional distributions of $\{x^\varepsilon(\cdot)\}$ to those of $x(\cdot)$ on an arbitrary time interval 0,T. Tightness (Theorem 4) is a little harder to get, owing to the fact that $f(x,\xi_t^\varepsilon)/\sqrt{\varepsilon}$ is not necessarily uniformly bounded as $\varepsilon \to 0$ (see remarks in Section 1), and some additional conditions must be used.

Theorem 3. Under (A1) to (A7), the finite dimensional distributions of $x^\varepsilon(\cdot)$ converge to those of $x(\cdot)$ as $\varepsilon \to 0$, where $x(\cdot)$ is the diffusion with infinitesimal operator A (on $\hat{C}^2$ functions) and initial condition x_0.

Remark. Give $k \in \hat{C}^3$, the main work in using Theorem 1 is in finding a suitable k^ε and proving (3.2)-(3.3). Following a basic idea in [5], [6], we will define functions k_1^ε and k_2^ε such that $k^\varepsilon = k + k_1^\varepsilon + k_2^\varepsilon$ does the job. The k_i^ε are constructed to guarantee that $\text{p-lim}\, k_i^\varepsilon = 0$ and $\hat{A}^\varepsilon k^\varepsilon = \hat{A}k$ plus terms going to zero in the sense of p-lim.

Proof: Part 1. Let $k \in \hat{C}^3$. Then it is easy to check that $k_0^\varepsilon(\cdot) \equiv k(x^\varepsilon(.)) \in \mathscr{D}(\hat{A}^\varepsilon)$ and that (write $x = x_t^\varepsilon$, $\xi = \xi_t^\varepsilon$)

$$\hat{A}^\varepsilon k_0^\varepsilon(t) = k_x'(x)[g(x,\xi) + f(x,\xi)/\varepsilon].$$

Part 2. We now define $k_1^\varepsilon(t) = k_1^\varepsilon(x_t,t)$ in such a way that $\hat{A}^\varepsilon k_1^\varepsilon(t)$ cancels the $k_x'(x)f(x,\xi)/\varepsilon$ term of $\hat{A}^\varepsilon k_0^\varepsilon(t)$. Define

$$k_1^\varepsilon(x,t) = \int_0^\infty E_t^\varepsilon k_x'(x)(f(x,\xi_{t+s}^\varepsilon)/\varepsilon)ds. \tag{5.4}$$

By (A3), (A4), ($i = [t/\varepsilon]$)

$$|k_1^\varepsilon(x,t)| \le \sum_{j=0}^\infty |E_i^\varepsilon k_x'(x)f(x,s_{i+j}^\varepsilon)| \le K \sum_j \mu_j. \tag{5.5}$$

Thus, k_1^ε is bounded. By (A4a), $E|E_i^\varepsilon k_x'(x)f(x,s_{i+j}^\varepsilon)| \le K\sqrt{\varepsilon}$. Thus (5.5) is bounded above by $K\varepsilon^{1/4}\sum_j \mu_j^{1/2}$ and, hence, $\text{p-lim}\; k_1^\varepsilon = 0$.

Note that $k_1^\varepsilon(x,t)$ is differentiable with respect to x. In fact,

$$(5.6) \qquad k_{1,x}^\varepsilon(x,t) = \int_0^\infty E_t^\varepsilon [k_x'(x)f(x,\xi_{t+s}^\varepsilon)/\varepsilon]_x ds$$

which is well defined for each ε. We can readily show that $\sup_t E|E_t^\varepsilon k_1(x_{t+\delta}^\varepsilon,\xi_{t+\delta}^\varepsilon) - k_1^\varepsilon(x_t^\varepsilon,\xi_t^\varepsilon)|/\delta$ is bounded as $\delta \to 0$ and that

$$\begin{aligned}
(5.7) \qquad & \text{p-lim}[E_t^\varepsilon k_1^\varepsilon(x_{t+\delta}^\varepsilon,t+\delta) - k_1^\varepsilon(x_t^\varepsilon,t)]/\delta \\
&= \text{p-lim}[E_t^\varepsilon k_1^\varepsilon(x_{t+\delta}^\varepsilon,t+\delta) - k_1^\varepsilon(x_t^\varepsilon,t+\delta)]/\delta \\
&+ \text{p-lim}[E_t^\varepsilon k_1^\varepsilon(x_t^\varepsilon,t+\delta) - k_1^\varepsilon(x_t^\varepsilon,t)]/\delta \\
&= (k_{1,x}^\varepsilon(x_t^\varepsilon,t))'\dot{x}_t^\varepsilon - k_x'(x_t^\varepsilon)f(x_t^\varepsilon,\xi_t^\varepsilon)/\delta \\
&\equiv \hat{k}_{1,x}(x_t^\varepsilon,t) + \tilde{k}(x_t^\varepsilon,t)/\varepsilon
\end{aligned}$$

where $\dot{x}_t^\varepsilon = g(x_t^\varepsilon,\xi_t^\varepsilon) + f(x_t^\varepsilon,\xi_t^\varepsilon)/\varepsilon$. Thus $k_1^\varepsilon \in \mathcal{D}(\hat{A}^\varepsilon)$ and $\hat{A}^\varepsilon k_1^\varepsilon$ is given by (5.7). The second term on the r.h.s. of (5.7) cancels a term of $\hat{A}^\varepsilon k_0^\varepsilon$. Also the term in $\hat{k}_{1,x}^\varepsilon(x_t^\varepsilon,t)'\dot{x}_t^\varepsilon$ which arises from the g term goes to zero in the p-lim sense as $\varepsilon \to 0$.

Part 3. Now k_2^ε is to be defined such that $\hat{A}^\varepsilon k_2^\varepsilon(t)$ equals $\hat{A}k(x_t^\varepsilon)$ minus the dominant term in $k_{1,x}^\varepsilon(x_t^\varepsilon,t)'\dot{x}_t^\varepsilon$ plus terms which go to zero in the p-lim sense as $\varepsilon \to 0$. Set $k_2^\varepsilon(t) = k_2^\varepsilon(x_t^\varepsilon,t)$ where

$$\begin{aligned}
(5.8) \qquad k_2^\varepsilon(x,t) &= \int_0^\infty ds \left\{\int_0^\infty E_t^\varepsilon \frac{f'(x,\xi_{t+s}^\varepsilon)}{\varepsilon}\left[k_x'(x)\,\frac{f(x,\xi_{t+s+u}^\varepsilon)}{\varepsilon}\right] du - A_{t+s}^\varepsilon k(x)\right\} \\
&+ \int_0^\infty E_t^\varepsilon k_x'(x)[g(x,\xi_{t+s}^\varepsilon) - \bar{g}^\varepsilon(x)]ds = \tilde{k}_2^\varepsilon(x,t) + \hat{k}_2^\varepsilon(x,t).
\end{aligned}$$

The last term of k_2^ε goes to zero uniformly in x, as $\varepsilon \to 0$; also (with $x = x_t^\varepsilon$) it is in $\mathcal{D}(\hat{A}^\varepsilon)$ and $\hat{A}^\varepsilon\hat{k}_2^\varepsilon(x,t) = k'(x)\bar{g}^\varepsilon(x)$ plus terms going to zero in the p-lim sense minus $k_x'(x)g(x,\xi_t^\varepsilon)$. This last term cancels a term of $\hat{A}^\varepsilon k_0^\varepsilon(t)$. We need to deal only with the first term of (5.8). Recall that $A_{t+s}^\varepsilon k(x)$ is just the expected value of the integral with respect to u of the coefficient of E_t^ε.

Owing to the integration with respect to s in (5.8) A^{ε}_{t+s} can (and will) be replaced by $\bar{A}^{\varepsilon}$ with little required change in the subsequent arguments. Next, we show that $\tilde{k}^{\varepsilon}_2$ is well-defined and p-lim $\tilde{k}^{\varepsilon}_2 = 0$. To simplify the notation, assume the scalar case and let f^{ε}_j represent either $f(x,s^{\varepsilon}_j)$ or $f_x(x,s^{\varepsilon}_j)$. All bounds and convergences are uniform on bounded x-sets.

We have $(i = [t/\varepsilon])$

$$|\tilde{k}^{\varepsilon}_2(t)| \leq K \sum_{i,j} |E^{\varepsilon}_i f^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k} - Ef^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k}|.$$

By (A4a,b) the i,j^{th} term is bounded above by $K\mu_j$ and also by $K\mu_k$, hence by $K(\mu_j\mu_k)^{1/2}$ which implies that the sum is bounded. Also, by (A4a,c), p-lim $\tilde{k}^{\varepsilon}_2 = 0$. It can also be readily shown that $\tilde{k}^{\varepsilon}_2 \in \mathscr{D}(\hat{A}^{\varepsilon})$. We need only show that $\hat{A}^{\varepsilon}\tilde{k}^{\varepsilon}_2$ = minus argument of the outer integral of $\tilde{k}^{\varepsilon}_2$ (at $s = 0$) plus $\tilde{k}^{\varepsilon}_{2,x}(x^{\varepsilon}_t,t)'\dot{x}^{\varepsilon}_t$ and that p-lim $\tilde{k}^{\varepsilon}_{2,x}(x^{\varepsilon}_t,t)'\dot{x}^{\varepsilon}_t = 0$. If this is true, then we have $\hat{A}^{\varepsilon}(k(x^{\varepsilon}_t) + k^{\varepsilon}_1(t) + k^{\varepsilon}_2(t)) = Ak(x^{\varepsilon}_t)$ plus terms whose p-lim is zero, and the proof will be completed. That $\hat{A}^{\varepsilon}\tilde{k}^{\varepsilon}_2$ has the asserted form is not hard to show, and we only show that p-lim $(\dot{x}^{\varepsilon})'\tilde{k}^{\varepsilon}_{2,x} = 0$.

To get $\tilde{k}^{\varepsilon}_{2,x}$, we can differentiate under the integral in (5.8). With the use of the above f^{ε}_i notation, $\tilde{k}^{\varepsilon}_{2,x}(t)'x^{\varepsilon}_t$ is bounded by a finite (the number not depending on ε) sum of terms of the type $(i = [t/\varepsilon])$

$$(5.9) \qquad Q^{\varepsilon} = |f^{\varepsilon}_i/\varepsilon| \sum_{i,j} |E^{\varepsilon}_i f^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k} - Ef^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k}|.$$

By (A4), $E^{1/2}|f^{\varepsilon}_i|^2 \leq K\varepsilon^{1/2}$ and $E^{1/2}|E^{\varepsilon}_i f^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k}|^2 \leq KEE^{\varepsilon}_i|f^{\varepsilon}_{i+j}|\mu_k \leq K\varepsilon^{1/2}\mu_k$. This and (A4c) implies that

$$E^{1/2}|E^{\varepsilon}_i f^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k} - Ef^{\varepsilon}_{i+j} f^{\varepsilon}_{i+j+k}|^2 \leq K\varepsilon^{(1+\delta)/2}(\mu_j\mu_k)^{1/2}.$$

Thus, by (A4) again, $E|Q^{\varepsilon}| \leq K\varepsilon^{\delta/2}$; hence, p-lim $(\dot{x}^{\varepsilon})'\tilde{k}^{\varepsilon}_{2,x} = 0$. Q.E.D.

Tightness. To prove tightness (which together with the convergence of finite dimensional distributions yields weak convergence of $\{x^{\varepsilon}(\cdot)\}$ to $x(\cdot)$, we make use of Theorem 2. Both (4.4) and (4.5) must be shown for each $k \in \hat{C}^3$. They do not follow from (A1)-(A7). In order to simplify matters we use (A8) also.

(A8) Let $s_i^\varepsilon = L_\varepsilon w_i^\varepsilon$, where L_ε is a matrix (bounded uniformly in ε) and $\{w_i^\varepsilon, i \geq 0\}$ is a Markov chain for each ε. There are matrices ρ_j^ε such that $\sum_j |\rho_j^\varepsilon| < \infty$ and where the convergence is uniform in ε, and $E_i^\varepsilon w_{i+j}^\varepsilon = \rho_j^\varepsilon w_i^\varepsilon$. We use E_i^ε to denote conditioning on $(w_j^\varepsilon, j \leq i)$. Let $E|w_i^\varepsilon|^7 \leq K\varepsilon^{7/2}$ and $|E_i^\varepsilon w_{i+j}^\varepsilon (w_{i+j}^\varepsilon)' - Ew_{i+j}^\varepsilon (w_{i+j}^\varepsilon)'| \leq \rho_j^\varepsilon(|w_i^\varepsilon|^2 + K\varepsilon)$.

Remark. Condition (A8) is quite realistic. For example, let $w_{n+1}^\varepsilon = Bw_n + \beta_n^\varepsilon$, where the eigenvalues of B are inside the unit circle and for each ε, $\{\beta_n^\varepsilon\}$ is a sequence of truncated independent identically distributed Gaussian random variables with covariance bounded above by $K\varepsilon$. Then (we use scalar case for illustration) $E_i^\varepsilon (w_{i+j}^\varepsilon)^2 = B^{2j}(w_i^\varepsilon)^2 + \sum_{k=0}^{j-1} (\text{var } \beta_n^\varepsilon) B^{2k}$ and $|E_i^\varepsilon (w_{i+j}^\varepsilon)^2 - E(w_{i+j}^\varepsilon)^2| \leq |B|^{2j}(w_i^\varepsilon)^2 + \sum_{k=j}^{\infty} (\text{var } \beta_n^\varepsilon)|B|^{2k}$ where $\text{var } \beta_n^\varepsilon \leq K\varepsilon$.

Theorem 4. Under (A1) to (A8), $\{x^\varepsilon(\cdot)\}$ is tight on $D^r[0,\infty)$ and converges weakly to the diffusion $x(\cdot)$ of Theorem 3 as $\varepsilon \to 0$.

Proof. Tightness of $\{k(x^\varepsilon(\cdot))\}$ on $D[0,T]$ needs to be shown, for arbitrary T and $k \in \hat{C}^3$. Define k_i^ε as in Theorem 3. Since k_i^ε, $i = 0,1,2$, are bounded, and $k^\varepsilon \in \mathcal{D}(\hat{A}^\varepsilon)$, then $(k^\varepsilon)^2 \in \mathcal{D}(\hat{A}^\varepsilon)$ and $\hat{A}^\varepsilon (k^\varepsilon)^2 = 2k^\varepsilon \hat{A}^\varepsilon k^\varepsilon$. By Theorem 2 and the estimates of Theorem 3 we only need to show that for each $\delta > 0$ $(i = 1,2)$ $P\{\sup_{t \leq T} |k_i^\varepsilon(t)| \geq \delta\}^2 \to 0$ as $\varepsilon \to 0$ and that $(\tilde{k}_{2,x}^\varepsilon)' \dot{x}^\varepsilon$ is bounded on $[0,T]$ by some random variable $M_\varepsilon(T)$ such that $\sup_\varepsilon P\{M_\varepsilon(T) \geq A\} \to 0$ as $A \to \infty$. (All the other components of $\hat{A}^\varepsilon k^\varepsilon$ have this property by the estimates in Theorem 3.)

By (A8) (using scalar case notation where convenient)

$$(5.10) \quad \sup_{t \leq T} |k_1^\varepsilon(t)| \leq K \sup_{i \leq [T/\varepsilon]} \sum_{j=0}^{\infty} |E_i^\varepsilon s_{i+j}^\varepsilon| \leq K \sup_{i \leq [T/\varepsilon]} |w_i^\varepsilon|$$

$$(5.11) \quad \sup_{t \leq T} |\tilde{k}_2^\varepsilon(t)| \leq K \sup_{i \leq [T/\varepsilon]} \sum_{j,k=0}^{\infty} |E_i^\varepsilon s_{i+j}^\varepsilon s_{i+j+k}^\varepsilon - E s_{i+j}^\varepsilon s_{i+j+k}^\varepsilon|$$

$$\leq K \sup_{i \leq [T/\varepsilon]} |w_i^\varepsilon|^2 + K\varepsilon.$$

Similarly

$$(5.12)\qquad \sup_{t\le T}|\tilde{k}^{\varepsilon}_{2,x}(t)'\dot{x}^{\varepsilon}_{t}| \le K\frac{|s^{\varepsilon}_{i}|}{\varepsilon}\sum_{j,k=0}^{\infty}|E^{\varepsilon}_{i}s^{\varepsilon}_{i+j}s^{\varepsilon}_{i+j+k} - Es^{\varepsilon}_{i+j}s^{\varepsilon}_{i+j+k}|$$

$$\le K \sup_{i\le[T/\varepsilon]}\frac{|w^{\varepsilon}_{i}|^{3}}{\varepsilon} + K_{0}|w^{\varepsilon}_{i}|.$$

By the comments in the first paragraph and the bounds (5.10)-(5.12), we need only show that

$$(5.13)\qquad P^{\varepsilon} \equiv P\{\sup_{i\le[T/\varepsilon]}|w^{\varepsilon}_{i}| \ge \Lambda\varepsilon^{1/3}\} \to 0$$

as $\varepsilon \to 0$ for each positive Λ. Bounding (5.13) and using Chebychev's inequality yields

$$P^{\varepsilon} \le \sum_{i=0}^{[T/\varepsilon]} P\{|w^{\varepsilon}_{i}| \ge \Lambda\varepsilon^{1/3}\} \le (\frac{T}{\varepsilon}+2)\frac{E|w^{\varepsilon}_{i}|^{7}}{\Lambda^{7}\varepsilon^{7/3}} \le \frac{K\varepsilon^{1/6}}{\Lambda^{7}},$$

from which we get that (5.10), (5.11) go to zero as desired and that (5.12) is bounded as desired. In fact (5.12) goes to zero in probability as $\varepsilon \to 0$. Q.E.D.

6. Convergence and Stability for the Adaptive Antenna Problem

We use the same notation as in Section 2. The symbols E^{ε}_{t} and E_t refer to expectation conditioned on $y(s)$, $s \le t/\varepsilon^2$ and $y(s)$, $s \le t$, resp. Write $K_0 = (K\bar{M}_0+I)/\tau$. As in Section 2, $y(\cdot)$ is a (not necessarily piecewise constant) stationary right continuous and bounded process. The quantities are not complex valued here, to simplify the notation. In general, deal with the real and imaginary parts separately. Assume

(B1) $\int_0^\infty |E_t\delta\tilde{M}_{t+s}|ds$ <u>is bounded uniformly in</u> t,ω.

Define the operator A^0_t (on $\hat{C}^2$ functions) by ($\bar{w}_t$ is not random)

$$A^0_t k(u) = \frac{K^2}{\tau^2}\int_0^\infty \bar{w}'_t E\delta\tilde{M}'_0 k_{uu}(u)\delta\tilde{M}_s\bar{w}_t ds.$$

(B2) $\int_0^\infty ds\int_0^\infty |E_t\delta\tilde{M}_{t+s}\delta\tilde{M}'_{t+s+v} - E\delta\tilde{M}_{t+s}\delta\tilde{M}'_{t+s+v}|dv$

<u>is bounded uniformly in</u> t,ω.

(B3) Either let E_t be the expectation conditioned on $\bigcap_{\rho>0} \mathscr{B}(y(s), s \le t + \rho)$ (and similarly for E_t^ε) or else $E_{t+\rho} y_{t+s} \to E_t y_{t+s}$ and $E_{t+\rho} y_{t+s} y'_{t+\tau} \to E_t y_{t+s} y'_{t+\tau}$ (and also for 3^{rd} order terms) in probability as $\rho \downarrow 0$, for each positive t, s, τ.

Both (B1) and (B2) follow from the strong mixing condition (1.3).

Define the operator A_t by $A_t k(u) = -k_u'(u) K_0 u + A_t^0 k(u) \equiv -k_u'(u) K_0 u + \frac{1}{2} \sum_{i,j} p_{ij}(t) \partial^2 k(u)/\partial u_i \partial u_j$. Set $\{p_{ij}\} = P = QQ'$.

Theorem 5. Under (B1)-(B3), $\{u^\varepsilon(\cdot)\}$ converges weakly in $D^r[0,\infty)$ to the diffusion (2.5) with generator A_t and initial condition $u(0) = 0$.

The proof follows the lines of that of Theorem 3 quite closely, except, of course, that the integrals cannot necessarily be written as sums. Also, due to the scaling and to the fact that $y(\cdot)$ is bounded, tightness is more easily proved than in Section 5. In particular, $k_i^\varepsilon \to 0$ as $\varepsilon \to 0$ $(i = 1,2)$ and so does $(k_{2,u}^\varepsilon)'\dot{u}^\varepsilon$, all uniformly on $[0,\infty)$. Theorem 5 is also a consequence of the results in [4] (under their mixing condition (3.1) on $y(\cdot)$, which implies (B1)-(B2)), [6] (for Markovian $y(\cdot)$) or [7]. We will give only the relevant k_i^ε.

Let $k \in \hat{C}^3$. Then $k_i^\varepsilon(t) = k_i^\varepsilon(u_t^\varepsilon, t)$, where we use

$$k_1^\varepsilon(u,t) = \frac{-K}{\tau} \int_0^\infty E_t^\varepsilon k_u'(u) \left[\delta\tilde{M}_{t+s}^\varepsilon u + \frac{\delta\tilde{M}_{t+s}}{\varepsilon} \bar{w}_t \right] ds$$

$$k_2(u,t) = \frac{K^2}{\tau^2} \int_0^\infty ds \left\{ \int_0^\infty \bar{w}_t' E_t^\varepsilon \frac{(\delta\tilde{M}_{t+s+v})'}{\varepsilon} k_{uu}(u) \frac{\delta\tilde{M}_{t+s}}{\varepsilon} \bar{w}_t' dv - A_t^0 k(u) \right\}.$$

We have $\hat{A}^\varepsilon k^\varepsilon = A_t k$ plus terms which to to zero (uniformly in ω, t) as $\varepsilon \to 0$.

An extension of Theorem 5. Theorem 5 is not very satisfactory, since we are usually interested in $u^\varepsilon(\cdot)$ for large times, and would like some estimates of how close the tails ($[T,\infty)$ sections, for large T) are to the tail of (2.5). The weak convergence result of Theorem 5 does not give this to us. In Theorem 6, we show that

(6.1) $$\sup_{t>0,\,\varepsilon \text{ small}} E|u^\varepsilon(t)|^2 \text{ is bounded}$$

(6.2) $\{u^\varepsilon(T+\cdot),\ T > 0,\ \varepsilon \text{ small}\}$ is tight in $D^r[0,\infty)$.

From these, we get the more satisfactory conclusion of Theorem 6. As noted after the proof, the method is not sharp enough to get good estimates for how small ε must be in order for (6.1)-(6.2) to hold.

Theorem 6. Assume (B1)-(B3). Then (6.1) and (6.2) hold and as $\varepsilon \to 0$, $T \to \infty$ in any way at all, $\bar{u}^\varepsilon(T+\cdot)$ converges weakly to the stationary solution to (2.5). Also, if $u^\varepsilon(T+\cdot)$ converges weakly to a process $u^\varepsilon(\cdot)$ as $T \to \infty$, then $u^\varepsilon(\cdot)$ converges weakly to the stationary solution of (2.5). (Indeed, it can be easily shown from (6.1) and the stationarity of $y(\cdot)$ that $\{u^\varepsilon(T+\cdot),\ T > 0\}$ is tight for each small ε.)

Proof: Only a sketch will be given. The stability idea is essentially that in [6], except that it is used in a slightly different way and that Kurtz's results must be used since $y(\cdot)$ is not assumed to be Markovian. Let P be a positive definite symmetric matrix such that $u'Pu \equiv k(u)$ is a Liapunov function for $\dot{u} = -K_0 u$, and $C = -(K_0'P+PK_0)$ is negative definite. In the proofs of Theorems 3, 4, 5, it was required that $k(\cdot) \in \hat{C}^3$. Our $k(\cdot)$ here is not in $\hat{C}^3$, but it makes little difference in the proofs. This is because $y(\cdot)$ is bounded and because we could use the form $\int_t^T E_t^\varepsilon(\cdot)$ for the k_i^ε and get estimates (6.1) for $t \le T$, and which do not depend on T. In order to simplify the argument, we ignore the fact that $k(\cdot) \notin \hat{C}^3$.

We have (use $u = u_t^\varepsilon$, $\bar{w} = \bar{w}_t$, $\delta\tilde{M}^\varepsilon = \delta\tilde{M}_t^\varepsilon$)

$$\hat{A}^\varepsilon k(u) = -u'Cu - \frac{K}{\tau} u'[(\delta\tilde{M}^\varepsilon)'P + P(\delta\tilde{M}^\varepsilon)]u - \frac{K}{\tau}[u'P\frac{\delta\tilde{M}^\varepsilon\bar{w}}{\varepsilon} + \bar{w}'(\frac{\delta\tilde{M}^\varepsilon}{\varepsilon})'Pu].$$

Following the method of Theorem 4, define k_1^ε such that $\hat{A}^\varepsilon k_1^\varepsilon$ cancels all the terms of $\hat{A}^\varepsilon k(u)$, except for the first. Set $k_1^\varepsilon(t) = k_1^\varepsilon(u_t^\varepsilon,t)$ where

(6.3) $$k_1^\varepsilon(u,t) = \frac{-K}{\tau}\int_0^\infty E_t^\varepsilon u'[(\delta\tilde{M}_{t+s}^\varepsilon)'P + P\delta\tilde{M}_{t+s}^\varepsilon u]ds$$

$$- \frac{K}{\tau\varepsilon}\int_0^\infty E_t^\varepsilon[u'P\delta\tilde{M}_{t+s}^\varepsilon\bar{w} + \bar{w}'(\delta\tilde{M}_{t+s}^\varepsilon)'Pu]ds.$$

Note that, by the change of variable $s/\varepsilon^2 \to s$,

$$(6.4)\quad k_1(u,t) = \frac{-K\varepsilon^2}{\tau}\int_0^\infty E_t^\varepsilon u'[\delta\tilde{M}(s+t/\varepsilon^2)'P + P\delta\tilde{M}(s+t/\varepsilon^2)]u\,ds$$

$$- \frac{K\varepsilon}{\tau}\int_0^\infty E_t^\varepsilon[u'P\delta\tilde{M}(s+t/\varepsilon^2)\overline{w} + \overline{w}'\delta\tilde{M}(s+t/\varepsilon^2)'Pu]ds.$$

Now,

$$(6.5)\qquad \hat{A}^\varepsilon k_1^\varepsilon(u,t) = -(\text{last two terms of } \hat{A}^\varepsilon k(u)) + (k_{1,u}^\varepsilon(u,t))'\dot{u}^\varepsilon.$$

The function $k_2^\varepsilon(t)$ is chosen in a way such that $\hat{A}^\varepsilon k_2^\varepsilon(t)$ cancels the part of the last term of (6.5) which is not $O(\varepsilon)$ (to see which terms are $O(\varepsilon)$ or $O(1)$, change variables $s/\varepsilon^2 \to s$). Thus set

$$(6.6)\qquad k_2^\varepsilon(t) = \frac{2K^2}{\tau^2\varepsilon^2}\int_0^\infty ds\int_0^\infty [E_t^\varepsilon\overline{w}_t'(\delta\tilde{M}_{t+s}^\varepsilon)'P\delta\tilde{M}_{t+s+v}^\varepsilon\overline{w}_t - C_t]dv$$

where

$$C_t = (2K^2/\tau^2)\int_0^\infty E\overline{w}_t'(\delta\tilde{M}_0)'P\delta\tilde{M}_s,\overline{w}_t ds.$$

Now,

$$\hat{A}^\varepsilon k_2^\varepsilon(t) = -(\text{dominant part of last term in (6.5)} + C_t.$$

By a change of variables $s/\varepsilon^2 \to s$, $v/\varepsilon^2 \to v$ and (B2), it can be verified that $k_2^\varepsilon = O(\varepsilon^2)$.

Combining the foregoing together with (B1)-(B2) yields, for $k^\varepsilon = k + k_1^\varepsilon + k_2^\varepsilon$, $u = u_t^\varepsilon$,

$$\hat{A}^\varepsilon k^\varepsilon(t) = -u'Cu + C_t + B_\varepsilon(u,\overline{w}_t)\varepsilon,$$
$$k(u) \le k^\varepsilon(u,t) + \varepsilon|\overline{B}_\varepsilon(u,\overline{w}_t)|,$$

where $B_\varepsilon(\cdot)$ and $\overline{B}_\varepsilon(\cdot)$ are both the sum of quadratic forms in u and bilinear forms in $(u,\overline{w}_t)$ and with coefficients that are bounded uniformly in ε,ω,t.

By (3.1) and the above estimates, there are positive real C,α and ε_0 such that for $\varepsilon \le \varepsilon_0$,

$$(6.7) \qquad Ek(u_s^\varepsilon) \le Ek^\varepsilon(u_0^\varepsilon) - \alpha \int_0^s Ek(u_v^\varepsilon)dv + \int_0^s C\, dv + \varepsilon|\overline{B}_\varepsilon(u_s^\varepsilon, \overline{w}_s)|.$$

Since $u_0^\varepsilon = 0$, (6.7) implies (6.1).

Given (6.1), it is not hard to show, via the method of Section 4 (see also Section 5 of [7]) that (6.2) holds. Also, the limit of any weakly convergent sequence (as $\varepsilon \to 0$) must converge to some solution of (2.5).

We only need to show that as $\varepsilon \to 0$, $T \to \infty$, any subsequence converges to the stationary solution of (2.5). Fix $T_1 > 0$ and take a convergent subsequence of $\{u^\varepsilon(T+\cdot)\}$ $(\varepsilon \to 0, T \to \infty)$. Take a further subsequence of the subsequence, such that the $[T-T_1,\infty)$ sections are weakly convergent also. Let $\overline{u}(\cdot)$ and $\overline{\overline{u}}(\cdot)$ denote the weak limits on $D^r[0,\infty)$ of the $[T,\infty)$ and $[T-T_1,\infty)$ sections, resp. Then $\overline{u}(\cdot)$ is just the $[T_1,\infty)$ section of $\overline{\overline{u}}(\cdot)$. In particular, $\overline{u}(0) = \overline{\overline{u}}(T_1)$. Since $\{u^\varepsilon(t)\}$ is tight and T_1 arbitrary and $\dot{u} = -K_0 u$ asymptotically stable, we get that any limit as $T \to \infty$, $\varepsilon \to 0$ must be the stationary solution of (2.5). Q.E.D.

Remark. Theorem 6 is preferable to Theorem 5, but since ε_0 depends on the maximum magnitude of $y(\cdot)$, we do not get a good estimate of the stability region. Some other approach seems to be needed for this. We have tried to combine the above ideas with the ideas in stochastic stability for linear systems with coefficient variations (such as those based on Gronwall's Lemma [15]-[17]) but without much success so far.

REFERENCES

[1] E. Wong and M. Zakai, "On the relationship between ordinary and stochastic differential equations", Int. J. Engin. Science, (1965), 3, 213-229.

[2] E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals", Ann. Math. Statist. 36, 1560-1564.

[3] R. Z. Khasminskii, "A limit theorem for solutions of differential equations with random right hand sides", Theory of Prob. and Applic., (1966), 11, 390-406.

[4] G. C. Papanicolaou and W. Kohler, "Asymptotic theory of mixing stochastic ordinary differential equations", Comm. Pure and Appl. Math., (1974), 27, 641-668.

[5] G. C. Papanicolaou, "Some probabilistic problems and methods in singular perturbations", Rocky Mountain Journal of Math., (1976), 6, 653-674.

[6] G. Blankenship and G. C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances", SIAM J. on Appl. Math., (1978), 34, 437-476.

[7] H. J. Kushner, "Jump diffusion approximations for ordinary differential equations with random right hand sides", submitted to SIAM J. on Control; see also LCDS Report 78-1, September 1978, Brown University.

[8] T. G. Kurtz, "Semigroups of conditional shifts and approximation of Markov processes", Ann. Prob., (1975), 4, 618-642.

[9] IEEE Trans. on Antennas and Propagation, (1976), AP-24. Special Issue on Adaptive Antenna Arrays.

[10] L. E. Brennan, E. L. Pugh and I. S. Reed, "Control-loop noise in adaptive antenna arrays", IEEE Trans. on Aerospace and Electronic Systems, (1971), AES-7, 254-262.

[11] I. I. Gikhman and A. V. Skorokhod, Introduction to the Theory of Random Processes, (1965), Saunders, Philadelphia.

[12] P. Billingsley, Convergence of Porbability Measures, (1968), John Wiley and Sons, New York.

[13] G. L. Blankenship and G. C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances", (1978), preprint.

[14] H. J. Kushner, Probability Methods for Approximations for Elliptic Equations and Optimal Stochastic Control Problems, Academic Press, New York, 1977.

[15] E. F. Infante, "On the stability of some linear autonomous random systems", (1968), ASME J. Appl. Mech., 35, 7-12.

[16] F. Kozin and C. M. Wu, "On the stability of linear stochastic differential equations", (1973), ASME J. Appl. Mech., 40, 87-92.

[17] G. Blankenship, "Stability of linear differential equations with random coefficients", (1977), IEEE Trans. on Automatic Control, AC-22, 834-838.